EEU44C04 / CS4031 / CS7NS3 / EEP55C27
Next Generation Networks

M/G/1 queue

Nicola Marchetti

nicola.marchetti@tcd.ie

Limitations of M/M/… to model telecommunication systems

- **M/M/… systems** are tractable due to the memoryless property of the interarrival and service times
- However, exponential *service* times may not be a good assumption, for example …
    - Some networks employ fixed packet sizes (deterministic service time)
    - There are limits on packet sizes, even when they are not fixed
- Poisson *arrival* assumption somewhat better due to aggregation of arrival streams
    - We will see later that this is also a problematic assumption in some networks

M/G/1 queue

- The **M/G/1 queue**, like the M/M/1 queue, has a Poisson arrival process, but it allows general distributions of service times
- Still assume that service times are …
  - Identically distributed
  - Mutually independent
  - Independent of interarrival times

Can we model M/G/1 as a Markov chain?

- We would need to include enough state information so that future state transitions depend *only* on the present state

- Number in the system, $n(t)$, is not enough

- Service time is not memoryless (not exponentially distributed anymore), so we need to know how long the current customer has been in service – call it $s_0(t)$
  - Would need to use state pairs { $n(t)$, $s_0(t)$ }

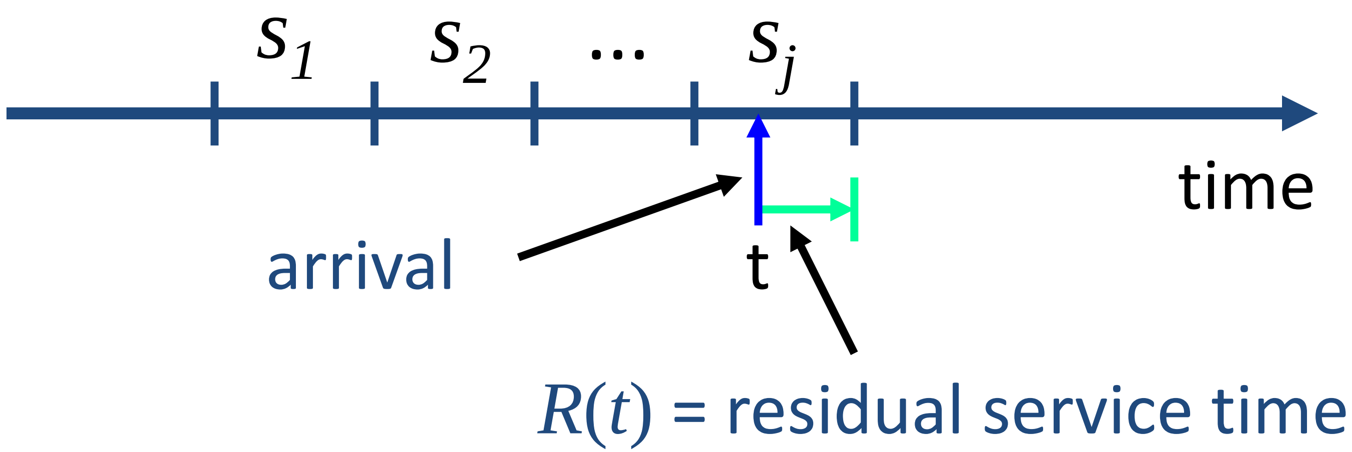
Analysis

- Significant problems affecting tractability
  - Two-dimensional state
  - $s_0(t)$ is a continuous state process, i.e., the state space is no longer discrete

- The **Pollaczek-Khinchin formula** (or just the "P-K formula") provides results for the M/G/1 queue
  - One analysis is based on residual service times
  - We'll just present results after a brief look at the residual service time

Residual service time

- Let s_1, s_2, … be the independent and identically distributed (i.i.d.) sequence of service times in an M/G/1 system
- When an arriving customer finds the server busy, the **residual service time** is the remaining service time for the customer now in service



Moments of the service time

- Let X_i be the service time of the i-th customer, e.g., the time needed to transmit the i-th packet
  - $\{X_1, X_2, \ldots\}$ are i.i.d. random variables and are independent of the interarrival times
  - Average service time

$$\overline{X} = E\{X\} = 1/\mu$$

  - Second moment of service time

$$\overline{X^2} = E\{X^2\}$$

P-K formula

- The Pollaczek-Khinchin formula gives the expected waiting time in queue for the M/G/1 queue

$$W = \frac{\lambda \overline{X^2}}{2(1-\rho)}$$

where ρ is the *utilization*:

Utilization: proportion of the system's resources used by the traffic which arrives to the system (should be < 1)

$$\rho = \lambda/\mu = \lambda \overline{X}$$

- Only the first and second moments of the service time distribution must be known!

Other M/G/1 results (1)

- The expected total time in the system, including queuing and service, is

$$T = \overline{X} + W = \overline{X} + \frac{\lambda \overline{X^2}}{2(1-\rho)}$$

- The P-K formula and Little's Law give the expected number of customers in queue, N_Q

$$N_Q = \lambda W = \lambda \left[\frac{\lambda \overline{X^2}}{2(1-\rho)} \right] = \frac{\lambda^2 \overline{X^2}}{2(1-\rho)}$$

Other M/G/1 results (2)

- The expected total number of customers in the system, N can also be determined using Little's Law

$$N = \lambda T = \lambda\left(\overline{X} + W\right)$$

$$= \lambda\overline{X} + \frac{\lambda^2\overline{X^2}}{2(1-\rho)} = \rho + \frac{\lambda^2\overline{X^2}}{2(1-\rho)}$$

$$= \rho + N_Q$$

utilization **customers in queue**

Problem

A queuing system has a Poisson arrival process and the service times are identically distributed, mutually independent and independent of interarrival times. The service is provided by a unique server. The average service time is 5 seconds and the arrival rate is 3/sec.

Write the Kendall's notation (motivating your choice) and calculate the utilization of the system. Is the system well dimensioned to properly serve its users? Why?

- M/G/1
  - Poisson arriv. → M
  - general distrib of service time → G
  - 1 server

$$\rho = \lambda \overline{X} = 3 \cdot 5 = 15$$

Being $\rho > 1$ the system is clearly not able to properly sustain incoming traffic. In fact the service rate is

$$\mu = 1/\overline{X} = 0.2/\sec$$

which is smaller than the arrival rate λ=3/sec.